

A NOVEL CHROMOSOME 21 GENE MARKER,
COMPOSITIONS AND METHODS USING SAME

ACKNOWLEDGEMENT

This invention was made in part with Government
5 support under Grant No. HD17449-11, from the National
Institutes of Child Health and Human Development. The
Government may have certain rights in this invention.

BACKGROUND OF THE INVENTION :

A major endeavor in molecular genetics has been
10 made in generating maps of the human genome. Human genome
mapping consists, generally, of ordering genomic DNA
fragments on their chromosomes using several methods, such
as fluorescent in situ hybridization (FISH), somatic cell
hybrid analysis or random clone fingerprinting. DNA
15 fragments that correspond to marked polymorphic sites can
be ordered by genetic linkage analysis. Distances between
polymorphic loci are estimated by meiotic recombination
frequencies. High resolution maps based upon the estimated
distances, however, cannot be constructed easily using such
20 methods because the resolution is low at the molecular
level and recombination frequency is not linearly
correlated with physical distance.

Thus, various obstacles such as, for example,
the difficulty in obtaining highly informative markers and
25 the paucity of identified markers that are evenly spaced
along the chromosome are significant weaknesses of the
currently available genetic maps. Most of the mapped
markers are restriction fragment length polymorphisms
(RFPLs) assayed by DNA hybridization. Although maps based
30 on these markers have contributed greatly to the primary
mapping of a number of diseases, they are still
insufficient for many applications such as mapping rare
monofactorial diseases, refining linkage intervals to

distances suited for gene identification, and mapping of loci contributing to complex traits.

Genetic linkage mapping is an important technology applied to the study of human biology and, in particular, for the delineation of the molecular basis of disease. Indeed, one of the most commonly used strategies for studying human inherited diseases is by cloning the responsible gene based on chromosomal location. Genetic linkage maps, therefore, facilitate the identification and mapping of genes involved in monogenic diseases, genes involved in multifactorial disorders, and are useful in carrier detection and prenatal diagnosis of hereditary disorders. A detailed linkage map is also a prerequisite for clone-based physical mapping and DNA sequencing of the entire chromosome.

Human chromosome 21 is a paradigm for large-scale human genome mapping efforts. The smallest human chromosome, chromosome 21 has approximately 50 megabases (Mb) of DNA. Less than 1% of the 2000 genes estimated to be on chromosome 21 are known. A high resolution map of chromosome is of particular interest because of its apparent role in familial Alzheimer disease (FAD), Down's syndrome (DS), amyotrophic lateral sclerosis (ALS), and Finnish progressive myoclonus epilepsy (PME). A gene defect responsible for FAD has been localized to chromosome 21 on the basis of genetic linkage to three pericentromeric loci. The gene encoding the precursor of the Alzheimer-associated amyloid β protein (APP), the principle component of the senile plaques and cerebrovascular amyloid deposits of Alzheimer disease (AD), has also been mapped to chromosome 21.

The process of developing such a long-range contig map involves the identification and localization of landmarks in cloned genetic fragments. When there are

enough landmarks for the size of the cloned fragments, contigs are formed, and the landmarks are simultaneously ordered. Currently, YACs, or yeast artificial chromosomes, are utilized for most mapping of the human genome. YACs permit cloning of fragments of \geq about 500 Kb. However, some difficulties have been encountered with the manipulation of YAC libraries. For example, in various YAC libraries, a fraction of the clones result from co-cloning events, i.e., they include in a single clone noncontiguous DNA fragments. A high percentage of YAC clones, particularly clones having high molecular weight inserts, are chimeric. Chimeric clones map to multiple sites on the chromosome and, thus, hamper the progress of mapping and analysis. Another problem endemic to YAC cloning is caused by DNA segments that are unclonable or unstable and tend to rearrange and delete.

Bacteria Artificial Chromosomes (BACs), provide an alternative to the YAC system. BACs mitigate the most problematic aspects of YACs such as, for example the high rate of chimerism and clonal instability. BACs are based on the *E. coli* single-copy plasmid F factor and are capable of faithful propagation of DNA fragments greater than about 300 Kb in size. BACs have a number of physical properties that make them amenable to physical mapping, including easy manipulation and an absence of chimerism. The lack of chimerism and the capacity to propagate large exogenous insert DNAs make the BACs excellent candidates for chromosome walking and the generation of contiguous physical maps.

The need for molecular description of chromosome 21 derives directly from the association with several human genetic diseases. A map of contiguous units (contigs) covering this chromosome will speed the identification of the cause of these diseases. Indeed, a detailed map would provide immediate access to the genomic segment, including

any pathological locus, as soon as it has been localized by genetic linkage or cytogenetic analysis.

Thus, a need exists for identifying, characterizing, and mapping the numerous genes that occupy
 5 loci on chromosome 21, which will expedite the rapid translation of high resolution chromosome maps into biological, medical and diagnostic applications. The present invention satisfies this need and provides related advantages as well.

10

SUMMARY OF THE INVENTION

The present invention provides isolated nucleic acids encoding human EHOC-1 protein and isolated receptor proteins encoded thereby. Further provided are vectors containing invention nucleic acids, probes that hybridize
 15 thereto, host cells transformed therewith, antisense oligonucleotides thereto and compositions containing, antibodies that specifically bind to invention polypeptides and compositions containing, as well as transgenic non-human mammals that express the invention protein.

20

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows a physical map for the consensus region for HPE1.

Figure 2 shows a physical map for the consensus region for EPM1 in relation to the consensus region for
 25 HPE1. The locations of YAC clones, BAC clones and EHOC-1 were indicated by thick bars.

Figure 3A shows the regions in which EHOC-1 has homologies to transmembrane proteins. Region 1 represents 29.4% identity in a 34 amino acid overlap with rat sodium
 30 channel protein III. Region 2 represents 20.4% identity in

a 103 amino acid overlap with phosphoglycerate transport system regulatory protein of *Salmonella typhimurium*. Region 3 represents 29.1% identity in a 55 amino acid overlap with pyrophosphate-energized vacuolar membrane proton pump of *Arabidopsis thaliana*. Region 4 represents 24.0% identity in a 50 amino acid overlap with myosin-like protein of *Saccharomyces cerevisiae*. Region 5 represents 17.9% identity in a 39 amino acid overlap with rabbit cardiac muscle ryanodine receptor. Region 6 represents 21.0% identity in a 62 amino acid overlap with rat cardiac muscle sodium channel protein alpha subunit. Region 7 represents 40.7% identity in a 27 amino acid overlap with rat skeletal muscle sodium channel protein alpha subunit. Region 8 represents a 30.3% identity in a 33 amino acid overlap with dystrophin cysteine-rich domain.

Figure 3B shows a comparison of genetic distances to the EPM1 locus in centiMorgans as computed by linkage disequilibrium studies. (Lehesjoki et al., *Hum. Mol. Genet.* 2:1229-1234 (1993)).

20

DETAILED DESCRIPTION OF THE INVENTION

Progressive myoclonus epilepsies (PMEs) are a heterogenous group of diseases which are characterized by myoclonus, epileptic seizures and progressive neurological deterioration including ataxia and dementia Berkovic et al., *New Engl. J. Med.* 315:296-305 (1986). PME of Unverricht-Lundborg type (EPM1) is an autosomal recessive disorder with frequent consanguinity in Finland and Mediterranean regions with the incidence of at least 1:20,000 in Finland. Genetic linkage analysis revealed that the locus for EPM1 is on chromosome 21q22.3 Malafosse et al., *Lancet* 339:1080-1081 (1992) and excluded Lafora disease from this region which is also a member of PME Lehesjoki et al., *Neurology* 42:1545-1550 (1992). Linkage

disequilibrium analysis made it possible to narrow down the candidate region to 300kb spanning the loci of PFKL, D21S25 and D21S154 Lehesjoki et al., *Hum. Mol. Genet.* 2:1229-1234 (1993); Lehesjoki et al., *Human Genetics* 93:668-674 (1994).

5 Autoimmune polyglandular disease type I (APECED) was also mapped to chromosome 21q22.3 by linkage disequilibrium analysis Aaltonen, J., et al., *Nature Genet.* 8:83-87 (1994). APECED is an autosomal recessive disease
10 resulting in a variable combination of failure of the parathyroid glands, adrenal cortex, gonads pancreatic β cells, thyroid gland and gastric parietal cells. Additional affects of APECED include alopecia, vitiligo, hepatitis, chronic mucocutaneous candidiasis, dystrophy of the dental enamel and nails and keratopathy. APCED usually
15 manifests itself in childhood, but tissue specific symptoms may appear throughout adulthood. The APCED locus maps within 500 kb of D21S49 and D21S171.

 Holoprosencephaly is characterized by impaired cleavage of the embryonic forebrain and incomplete mid-
20 facial development that manifest as a wide range of midfacial anomalies including cyclopia, ethmocephaly cebocephaly, premaxillary agenesis, hypotelorism, and a single maxillary central incisor. The most commonly associated chromosomal abnormality includes dup(3p),
25 del(7q), deletions of chromosome 13, trisomy 13, trisomy 18, and triploidy (Munke, *AM J Med Genet* 34:237-245 (1989)). The etiology is heterogeneous and may include aneuploidies for chromosomes 2, 3, 7, 13, 18 and 21. In order to narrow down the candidate region for HPE1, the
30 deletion of 21(q22.3) was characterized in two HP patients by fluorescence *in situ* hybridization and quantitative Southern blot dosage analysis. For the smaller deletion, the regions for D21S25, D21S154, D21S171 and D21S44 were deleted and for D21S42 and D21S49 were not. Combining
35 these data with previous reports of deletion of 21q22.3

(D21S112-ter) without the holoprosencephaly phenotype indicate that the region responsible for holoprosencephaly spans the 1-2Mb region including PFKL and ITGB2 (CD18).

Four cases of holoprosencephaly with chromosome 21 anomalies have been published. Estabrooks et al. describe a minute deletion of chromosome 21(q22.3) (Estabrooks et al., AM J Med Genet, 36:306-309 (1990)) suggesting this region as a locus for holoprosencephaly (HPE1).

Described in the instant specification is the construction of the BAC (Bacterial Artificial Chromosome) Shizuya et al., *Proc. Natl. Acad. Sci. USA* 89:8794-8797 (1992) contig of this EPM1-APECED-HPE1 candidate region and the isolation of a novel gene from this contiguous map unit using a direct cDNA selection technique.

In order to isolate genes responsible for these diseases, a cDNA library from a 14-week trisomy 21 fetal brain was constructed using Uni-Zap XR (Stratagene, La Jolla, CA). More than 95% clones have inserts ranging from 1-4kb (avg. 2kb). In addition, a direct cDNA selection method was applied to BACs (Bacterial Artificial Chromosomes) in the 21q22.3 region. Using cDNA synthesized from trisomy 21 fetal brain, Sau3AI linkers were attached, the cDNA then was digested with Sau3AI, followed by attachment of a second pair of linkers and hybridized to biotinylated BAC DNAs which cover the candidate region. cDNA/BAC DNA hybrid molecules were captured on streptavidin coated magnetic beads, non-specific cDNA were washed out, and specifically hybridized cDNA were eluted and amplified by PCR. Twice selected PCR products were subcloned and analyzed. Southern blot analysis revealed that 21 out of 30 (70%) of fragments yielded unique bands of the original BACs. Using these fragments as probes, cDNAs (3kb, 4kb and 5kb) were isolated from the library. The 5kb cDNA subclone (EHOC-1) maps proximal to but neighboring D21S25 and showed homologies to transmembrane genes. The loci of these genes

all map within the consensus region where holoprosencephaly, EPM1 and APECED are localized.

DNA sequence analysis of 5kb cDNA showed a complete coding sequence of 3570 bp which revealed to have
 5 homologies to transmembrane proteins including three kinds of sodium channel proteins on amino acid sequence level. (SEQ ID NOS: 1-2; Figure 3).

Five types of BAC clone were isolated from the total human genomic DNA BAC library Shizuya et al., *Proc. Natl. Acad. Sci. USA* 89:8794-8797 (1992) by PCR screening
 10 method using STSs containing PFKL, D21S25, D21S154 and CD18. Physical maps of the HPE1-EPM1-APECED consensus region with these BAC clones and YAC clones Chumakov et al., *Nature* 359:380-387 (1992) is described in Figures 1
 15 and 2. BAC-1 (230kb) and BAC-2 (210kb) were positive for D21S25. BAC-3 (170kb) was positive for D21S25 and PFKL. Agarose gel electrophoresis of EcoRI-digested BAC DNAs and Southern blot analysis showed that these 3 BACs overlap
 20 each other. BAC-4 was identical to BAC-3. BAC-5 (100kb) was positive for CD18.

Direct cDNA selection was performed on 5 BAC DNAs (four of which were overlapping) which span the consensus region. EcoRI digestion of subclone DNAs revealed that 10% clones were chimeric. The average sizes of the inserts of
 25 non-chimeric clones were 400bp. Forty non-chimeric subclones of selected cDNAs were analyzed by using EcoRI-digested BAC DNA Southern blots. Twenty-eight clones (70%) showed unique signals on the BAC blots, 6 clones (15%) showed repetitive, and 6 clones (15%) did not show any
 30 signal on these blots. Using insert DNAs of these subclones as probes, a trisomy 21 fetal brain cDNA library was screened. Three overlapping cDNAs (3kb, 4kb and 5kb) containing poly (A) tail were isolated and designated EHOC-1.

The three overlapping EHOC-1 cDNA subclones were used for Southern blot analysis using EcoRI-digested BAC DNA blots. Only BAC-1 showed unique multiple band signals indicating that these cDNAs originated from BAC-1. Identical sizes of the signal bands indicated that these clones overlap each other. Complete sequence of the EHOC-1 5kb cDNA clone and partial sequence analysis of 3kb and 4kb clones showed that entire sequence of the 3kb clone and part of the sequence of 4kb clone are contained in the 5kb clone, but the 3' end of the 4kb clone was different from that of 5kb clone indicating the existence of splice variants of EHOC-1 cDNAs. Northern blot analysis using the insert of 5kb EHOC-1 cDNA revealed three transcripts (5.3kb, 7.5kb and 8kb) on multiple adult tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas). Fluorescence in-situ hybridization was also done on lymphocytes of a normal individual using insert of the EHOC-1 5kb cDNA subclone. Discrete signals were seen on chromosome 21q22.3 confirming the loci. The complete sequence of 5kb cDNA clone revealed an open reading frame of 3570bp (1190 amino acid). The initiator ATG was located within a good Kozak consensus sequence Kozak, M., *J. Mol. Biol.* 196:947-950 (1987); Kozak, M., *Nuc. Acid Res.* 15:8125-8148 (1987). Homology search of an amino acid sequence of this ORF to genes registered in Genbank/EMBL showed that this gene product has homologies to multiple transmembrane proteins including three types of sodium channel proteins (Figure 3).

Some neurologic disorders in humans are known to result from mutations in sodium channels Ptacek et al., *Cell* 67:1021-1027 (1991); Rojas et al., *Nature* 354:387-389 (1991); McClatchey et al., *Cell* 68:769-774 (1992); Ptacek et al., *Neuron* 8:891-897 (1992), calcium channels Ptacek et al., *Cell* 77:863-868 (1994); Jurkut-Rott et al., *Hum. Mol. Genet.* 3:1415-1419 (1994), and a potassium channel Browne et al., *Nature Genet.* 3:136-140 (1994). By using BLAST

computer program Altschul, S.J., et al., *J. Mol. Biol.* 215:403-410 (1994), one fibronectin domain (CxV YxC) was found at 356 < - > 401 a.a. The analysis also showed that the motif (Sxxx(I,L)E) occurs at 462, 670, 708, 716, 730, and 1078. This motif was searched for various protein databases and there were very few where it was present three or more times. These include; rat cartilage specific proteoglycan core protein, myosin, *drosophila* sevenless (4 copies), *drosophila* prospero (4 copies), and *drosophila* serendipity (3 copies). The latter three are mutants in development. Sevenless causes an eye defect, prospero defects in axon pathfinding, and serendipity defects in cellularization. It is reasonable that a defect in axonal routing may correlate with the phenotype of EHOC-1. The region beginning at 777 also has some homologies to multiple drug resistance genes and to the *drosophila* rutabaga gene. Rutabaga is involved in learning in *drosophila*.

Accordingly, the present invention provides isolated nucleic acids encoding a novel gene, EHOC-1, which exists in human chromosome 21, specifically at the q23.2 locus, which is the site of mutation(s) that cause PME, HPE1, and APECED. The term "nucleic acids" (also referred to as polynucleotides) encompasses RNA as well as single and double-stranded DNA and cDNA. As used herein, the phrase "isolated" means a nucleic acid that is in a form that does not occur in nature. One means of isolating a nucleic acid encoding an EHOC-1 polypeptide is to probe a mammalian genomic library with a natural or artificially designed DNA probe using methods well known in the art. DNA probes derived from the EHOC-1 gene are particularly useful for this purpose. DNA and cDNA molecules that encode EHOC-1 polypeptides can be used to obtain complementary genomic DNA, cDNA or RNA from human, mammalian, or other animal sources, or to isolate related cDNA or genomic clones by the screening of cDNA or genomic

libraries, by methods described in more detail below. Examples of nucleic acids are RNA, cDNA, or isolated genomic DNA encoding an EHOC-1 polypeptide. Such nucleic acids may have coding sequences substantially the same as the coding sequence shown in SEQ ID NO: 2. This invention also encompasses nucleic acids which differ from the nucleic acids shown in SEQ ID NO: 1, but which have the same phenotype, i.e., encode substantially the same amino acid sequence set forth in SEQ ID NO: 2.

Phenotypically similar nucleic acids are also referred to as "functionally equivalent nucleic acids". As used herein, the phrase "functionally equivalent nucleic acids" encompasses nucleic acids characterized by slight and non-consequential sequence variations that will function in substantially the same manner to produce the same protein product(s) as the nucleic acids disclosed herein. In particular, functionally equivalent nucleic acids encode polypeptides that are the same as those disclosed herein or that have conservative amino acid variations. For example, conservative variations include substitution of a non-polar residue with another non-polar residue, or substitution of a charged residue with a similarly charged residue. These variations include those recognized by skilled artisans as those that do not substantially alter the tertiary structure of the protein.

Further provided are nucleic acids encoding EHOC-1 polypeptides that, by virtue of the degeneracy of the genetic code, do not necessarily hybridize to the invention nucleic acids under specified hybridization conditions. Preferred nucleic acids encoding the invention polypeptide are comprised of nucleotides that encode substantially the same amino acid sequence set forth in SEQ ID NO: 2. Alternatively, preferred nucleic acids encoding the invention polypeptide(s) hybridize under high stringency conditions to substantially the entire sequence, or

substantial portions (i.e., typically at least 15-30 nucleotides) of the nucleic acid sequence set forth in SEQ ID NO: 1.

Stringency of hybridization, as used herein, refers to conditions under which polynucleotide hybrids are stable. As known to those of skill in the art, the stability of hybrids is a function of sodium ion concentration and temperature. (See, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual 2d Ed. (Cold Spring Harbor Laboratory, 1989; incorporated herein by reference).

Also provided are isolated peptides, polypeptides(s) and/or protein(s) encoded by the invention nucleic acids which are EHOC-1 polypeptides. The EHOC-1 polypeptide comprises a protein of approximately 1190 amino acids in length. The predicted amino acid sequence encoding the EHOC-1 polypeptide is set forth in SEQ ID NO: 2.

As used herein, the term "isolated" means a protein molecule free of cellular components and/or contaminants normally associated with a native *in vivo* environment. Invention polypeptides and/or proteins include any natural occurring allelic variant, as well as recombinant forms thereof. The EHOC-1 polypeptides can be isolated using various methods well known to a person of skill in the art. The methods available for the isolation and purification of invention proteins include, precipitation, gel filtration, ion-exchange, reverse-phase and affinity chromatography. Other well-known methods are described in Deutscher et al., Guide to Protein Purification: Methods in Enzymology Vol. 182, (Academic Press, 1990), which is incorporated herein by reference. Alternatively, the isolated polypeptides of the present invention can be obtained using well-known recombinant

methods as described, for example, in Sambrook et al., *supra.*, 1989).

5 An example of the means for preparing the invention polypeptide(s) is to express nucleic acids encoding the EHOC-1 in a suitable host cell, such as a bacterial cell, a yeast cell, an amphibian cell (i.e., oocyte), or a mammalian cell, using methods well known in the art, and recovering the expressed polypeptide, again using well-known methods. Invention polypeptides can be
10 isolated directly from cells that have been transformed with expression vectors, described below in more detail. The invention polypeptide, biologically active fragments, and functional equivalents thereof can also be produced by chemical synthesis. As used herein, "biologically active
15 fragment" refers to any portion of the polypeptide represented by the amino acid sequence in SEQ ID NO: 2 that can assemble into a cationic channel permeable to Ca^{2+} which is activated by acetylcholine. Synthetic polypeptides can be produced using Applied Biosystems, Inc. Model 430A or
20 431A automatic peptide synthesizer (Foster City, CA) employing the chemistry provided by the manufacturer.

As used herein, the phrase "EHOC-1" refers to recombinantly expressed/produced (i.e., isolated or substantially pure) proteins that contain highly
25 hydrophobic regions which predict potential membrane spanning regions, having homologies to multiple transmembrane proteins, including sodium channel, calcium channel and potassium channel proteins including variants thereof encoded by mRNA generated by alternative splicing
30 of a primary transcript, and further including fragments thereof which retain one or more of the aforementioned properties. As used herein, the phrase "functional polypeptide" means that binding of ligands, for example, cause transcriptional activation of EHOC-1 proteins. More

specifically, agonist activation of a "functional invention polypeptide" induces the protein to generate a signal.

Modification of the invention nucleic acids, polypeptides or proteins with the following phrases:

5 "recombinantly expressed/produced", "isolated", or "substantially pure", encompasses nucleic acids, peptides, polypeptides or proteins that have been produced in such form by the hand of man, and are thus separated from their native *in vivo* cellular environment. As a result of this

10 human intervention, the recombinant nucleic acids, polypeptides and proteins of the invention are useful in ways that the corresponding naturally occurring molecules are not, such as identification of selective drugs or compounds.

15 Sequences having "substantial sequence homology" are intended to refer to nucleotide sequences that share at least about 90% identity with invention nucleic acids; and amino acid sequences that typically share at least about 95% amino acid identity with invention polypeptides. It is

20 recognized, however, that polypeptides or nucleic acids containing less than the above-described levels of homology arising as splice variants or that are modified by conservative amino acid substitutions, or by substitution of degenerate codons are also encompassed within the scope

25 of the present invention.

The present invention provides the isolated polynucleotide operatively linked to a promoter of RNA transcription, as well as other regulatory sequences. As used herein, the phrase "operatively linked" refers to the

30 functional relationship of the polynucleotide with regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences. For example, operative linkage of a polynucleotide to a promoter refers

to the physical and functional relationship between the polynucleotide and the promoter such that transcription of DNA is initiated from the promoter by an RNA polymerase that specifically recognizes and binds to the promoter, and
5 wherein the promoter directs the transcription of RNA from the polynucleotide.

Promoter regions include specific sequences that are sufficient for RNA polymerase recognition, binding and transcription initiation. Additionally, promoter regions
10 include sequences that modulate the recognition, binding and transcription initiation activity of RNA polymerase. Such sequences may be *cis* acting or may be responsive to *trans* acting factors. Depending upon the nature of the regulation, promoters may be constitutive or regulated.
15 Examples of promoters are SP6, T4, T7, SV40 early promoter, cytomegalovirus (CMV) promoter, mouse mammary tumor virus (MMTV) steroid-inducible promoter, Moloney murine leukemia virus (MMLV) promoter, and the like.

Vectors that contain both a promoter and a
20 cloning site into which a polynucleotide can be operatively linked are well known in the art. Such vectors are capable of transcribing RNA *in vitro* or *in vivo*, and are commercially available from sources such as Stratagene (La Jolla, CA) and Promega Biotech (Madison, WI). In order to
25 optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' and/or 3' untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation codons or other sequences that may interfere with or reduce
30 expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites can be inserted immediately 5' of the start codon to enhance expression. (See, for example, Kozak, J. Biol. Chem. 266:19867 (1991)). Similarly, alternative codons,

encoding the same amino acid, can be substituted for coding sequences of the EHOC-1 polypeptide in order to enhance transcription (e.g., the codon preference of the host cell can be adopted, the presence of G-C rich domains can be reduced, and the like).

Also provided are vectors comprising the invention nucleic acids. Examples of vectors are viruses, such as baculoviruses and retroviruses, bacteriophages, cosmids, plasmids and other recombination vehicles typically used in the art. Polynucleotides are inserted into vector genomes using methods well known in the art. For example, insert and vector DNA can be contacted, under suitable conditions, with a restriction enzyme to create complementary ends on each molecule that can pair with each other and be joined together with a ligase. Alternatively, synthetic nucleic acid linkers can be ligated to the termini of restricted polynucleotide. These synthetic linkers contain nucleic acid sequences that correspond to a particular restriction site in the vector DNA. Additionally, an oligonucleotide containing a termination codon and an appropriate restriction site can be ligated for insertion into a vector containing, for example, some or all of the following: a selectable marker gene, such as the neomycin gene for selection of stable or transient transfectants in mammalian cells; enhancer/promoter sequences from the immediate early gene of human CMV for high levels of transcription; transcription termination and RNA processing signals from SV40 for mRNA stability; SV40 polyoma origins of replication and ColE1 for proper episomal replication; versatile multiple cloning sites; and T7 and SP6 RNA promoters for *in vitro* transcription of sense and antisense RNA. Other means are well known and available in the art.

Also provided are vectors comprising a nucleic acids encoding an EHOC-1 polypeptide, adapted for

expression in a bacterial cell, a yeast cell, an amphibian cell (i.e., oocyte), a mammalian cell and other animal cells. The vectors additionally comprise the regulatory elements necessary for expression of the nucleic acid in
5 the bacterial, yeast, amphibian, mammalian or animal cells so located relative to the nucleic acid encoding EHOC-1 polypeptide as to permit expression thereof. As used herein, "expression" refers to the process by which nucleic acids are transcribed into mRNA and translated into
10 peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may include splicing of the mRNA, if an appropriate eucaryotic host is selected. Regulatory elements required for expression include promoter sequences to bind RNA polymerase and
15 transcription initiation sequences for ribosome binding. For example, a bacterial expression vector includes a promoter such as the lac promoter and for transcription initiation the Shine-Dalgarno sequence and the start codon AUG (Sambrook et al. *supra*). Similarly, a eucaryotic
20 expression vector includes a heterologous or homologous promoter for RNA polymerase II, a downstream polyadenylation signal, the start codon AUG, and a termination codon for detachment of the ribosome. Such vectors can be obtained commercially or assembled by the
25 sequences described in methods well known in the art, for example, the methods described above for constructing vectors in general. Expression vectors are useful to produce cells that express the invention polypeptide.

This invention provides a transformed host cell
30 that recombinantly expresses the EHOC-1 polypeptide. The host cell has been transformed with a nucleic acid encoding a EHOC-1 polypeptide. An example is a mammalian cell comprising a plasmid adapted for expression in a mammalian cell. The plasmid contains a nucleic acid encoding an
35 EHOC-1 polypeptide and the regulatory elements necessary for expression of the invention protein. Various mammalian

cells may be utilized as hosts, including, for example, mouse fibroblast cell NIH3T3, CHO cells, HeLa cells, Ltk-cells, etc. Expression plasmids such as those described *supra* can be used to transfect mammalian cells by methods well known in the art such as calcium phosphate precipitation, DEAE-dextran, electroporation, microinjection or lipofection.

EHOC-1 polypeptides expressed recombinantly on eucaryotic cell surfaces may contain at least one EHOC-1 polypeptide, or may contain a mixture of peptides encoded by the host cell and/or subunits encoded by heterologous nucleic acids.

The present invention provides nucleic acid probes comprising nucleotide sequences capable of specifically hybridizing with sequences included within the nucleic acid sequence encoding an EHOC-1 polypeptide, for example, a coding sequence included within the nucleotide sequence shown in SEQ ID NO: 1. As used herein, a "probe" is a single-stranded DNA or RNA that has a sequence of nucleotides that includes at least 15 contiguous bases set forth in SEQ ID NO: 1. Preferred regions from which to construct probes include 5' and/or 3' coding sequences, sequences within the ORF, sequences predicted to encode transmembrane domains, sequences predicted to encode cytoplasmic loops, signal sequences, ligand binding sites, and the like. Full-length or fragments of cDNA clones can also be used as probes for the detection and isolation of related genes. When fragments are used as probes, preferably the cDNA sequences will be from the carboxyl end-encoding portion of the cDNA, and most preferably will include predicted transmembrane domain-encoding portions of the cDNA sequence. Transmembrane domain regions can be predicted based on hydropathy analysis of the deduced amino acid sequence using, for example, the method of Kyte and Doolittle, J. Mol. Biol. 157:105 (1982).

As used herein, the phrase "specifically hybridizing" encompasses the ability of a polynucleotide to recognize a sequence of nucleic acids that are complementary thereto and to form double-helical segments via hydrogen bonding between complementary base pairs. Nucleic acid probe technology is well known to those skilled in the art who will readily appreciate that such probes may vary greatly in length and may be labeled with a detectable agent, such as a radioisotope, a fluorescent dye, and the like, to facilitate detection of the probe. Invention probes are useful to detect the presence of nucleic acids encoding the EHOC-1 polypeptide. For example, the probes can be used for *in situ* hybridizations in order to locate biological tissues in which the invention gene is expressed. Additionally, synthesized oligonucleotides complementary to the nucleic acids of a nucleotide sequence encoding EHOC-1 polypeptide are useful as probes for detecting the invention genes, their associated mRNA, or for the isolation of related genes using homology screening of genomic or cDNA libraries, or by using amplification techniques well known to one of skill in the art.

Also provided are antisense oligonucleotides having a sequence capable of binding specifically with any portion of an mRNA that encodes the EHOC-1 polypeptide so as to prevent translation of the mRNA. The antisense oligonucleotide may have a sequence capable of binding specifically with any portion of the sequence of the cDNA encoding the EHOC-1 polypeptide. As used herein, the phrase "binding specifically" encompasses the ability of a nucleic acid sequence to recognize a complementary nucleic acid sequence and to form double-helical segments therewith via the formation of hydrogen bonds between the complementary base pairs. An example of an antisense oligonucleotide is an antisense oligonucleotide comprising chemical analogs of nucleotides.

Compositions comprising an amount of the antisense oligonucleotide, described above, effective to reduce expression of the EHOC-1 polypeptide by passing through a cell membrane and binding specifically with mRNA encoding a EHOC-1 polypeptide so as to prevent its translation and an acceptable hydrophobic carrier capable of passing through a cell membrane are also provided herein. The acceptable hydrophobic carrier capable of passing through cell membranes may also comprise a structure which binds to a receptor specific for a selected cell type and is thereby taken up by cells of the selected cell type. The structure may be part of a protein known to bind to a cell-type specific receptor.

Antisense oligonucleotide compositions

inhibit translation of mRNA encoding the invention polypeptides. Synthetic oligonucleotides, or other antisense chemical structures are designed to bind to mRNA encoding the EHOC-1 polypeptides and inhibit translation of mRNA and are useful as compositions to inhibit expression of EHOC-1 associated genes in a tissue sample or in a subject.

This invention provides a means to modulate levels of expression of EHOC-1 polypeptides by the use of a synthetic antisense oligonucleotide composition (hereinafter SAOC) which inhibits translation of mRNA encoding these polypeptides. Synthetic oligonucleotides, or other antisense chemical structures designed to recognize and selectively bind to mRNA, are constructed to be complementary to portions of the nucleotide sequences shown in SEQ ID NO: 1 of DNA, RNA or chemically modified, artificial nucleic acids. The SAOC is designed to be stable in the blood stream for administration to a subject by injection, or in laboratory cell culture conditions. The SAOC is designed to be capable of passing through the cell membrane in order to enter the cytoplasm of the cell

by virtue of physical and chemical properties of the SAOC which render it capable of passing through cell membranes, for example, by designing small, hydrophobic SAOC chemical structures, or by virtue of specific transport systems in the cell which recognize and transport the SAOC into the cell. In addition, the SAOC can be designed for administration only to certain selected cell populations by targeting the SAOC to be recognized by specific cellular uptake mechanisms which bind and take up the SAOC only within select cell populations. For example, the SAOC may be designed to bind to a receptor found only in a certain cell type, as discussed supra. The SAOC is also designed to recognize and selectively bind to the target mRNA sequence, which may correspond to a sequence contained within the sequence shown in SEQ ID NO: 1. The SAOC is designed to inactivate the target mRNA sequence by either binding to the target mRNA and inducing degradation of the mRNA by, for example, RNase I digestion, or inhibiting translation of the mRNA target by interfering with the binding of translation-regulating factors or ribosomes, or inclusion of other chemical structures, such as ribozyme sequences or reactive chemical groups which either degrade or chemically modify the target mRNA. SAOCs have been shown to be capable of such properties when directed against mRNA targets (see Cohen et al., TIPS, 10:435 (1989) and Weintraub, Sci. American, January (1990), pp.40; both incorporated herein by reference).

This invention provides a composition containing an acceptable carrier and any of an isolated, purified EHOC-1 polypeptide, an active fragment thereof, or a purified, mature protein and active fragments thereof, alone or in combination with each other. These polypeptides or proteins can be recombinantly derived, chemically synthesized or purified from native sources. As used herein, the term "acceptable carrier" encompasses any of the standard pharmaceutical carriers, such as phosphate

buffered saline solution, water and emulsions such as an oil/water or water/oil emulsion, and various types of wetting agents.

Also provided are antibodies having specific
5 reactivity with the EHOC-1 polypeptides of the subject invention. Active fragments of antibodies are encompassed within the definition of "antibody".

The antibodies of the invention can be produced by methods known in the art using the invention
10 polypeptides, proteins or portions thereof as antigens. For example, polyclonal and monoclonal antibodies can be produced by methods well known in the art, as described, for example, in Harlow and Lane, Antibodies: A Laboratory Manual (Cold Spring Harbor Laboratory 1988), which is
15 incorporated herein by reference. The polypeptide of the present invention can be used as the immunogen in generating such antibodies. Alternatively, synthetic peptides can be prepared (using commercially available synthesizers) and used as immunogens. Amino acid sequences
20 can be analyzed by methods well known in the art to determine whether they encode hydrophobic or hydrophilic domains of the corresponding polypeptide. Altered antibodies such as chimeric, humanized, CDR-grafted or bifunctional antibodies can also be produced by methods
25 well known in the art. Such antibodies can also be produced by hybridoma, chemical synthesis or recombinant methods described, for example, in Sambrook et al., *supra.*, and Harlow and Lane, *supra.* Both anti-peptide and anti-fusion protein antibodies can be used. (see, for example,
30 Bahouth et al., Trends Pharmacol. Sci. 12:338 (1991); Ausubel et al., Current Protocols in Molecular Biology (John Wiley and Sons, NY 1989) which are incorporated herein by reference).

The invention antibodies can be used to isolate the invention polypeptides. Additionally the antibodies are useful for detecting the presence of the invention polypeptides, as well as analysis of chromosome localization, and structure of functional domains. Methods for detecting the presence of an EHOC-1 polypeptide on the surface of a cell comprise contacting the cell with an antibody that specifically binds to the EHOC-1 polypeptide, under conditions permitting binding of the antibody to the polypeptides, detecting the presence of the antibody bound to the cell, and thereby detecting the presence of the invention polypeptide on the surface of the cell. With respect to the detection of such polypeptides, the antibodies can be used for *in vitro* diagnostic or *in vivo* imaging methods.

Immunological procedures useful for *in vitro* detection of the target EHOC-1 polypeptide in a sample include immunoassays that employ a detectable antibody. Such immunoassays include, for example, ELISA, Pandex microfluorimetric assay, agglutination assays, flow cytometry, serum diagnostic assays and immunohistochemical staining procedures which are well known in the art. An antibody can be made detectable by various means well known in the art. For example, a detectable marker can be directly or indirectly attached to the antibody. Useful markers include, for example, radionucleotides, enzymes, fluorogens, chromogens and chemiluminescent labels.

Further, invention antibodies can be used to modulate the activity of the EHOC-1 polypeptide in living animals, in humans, or in biological tissues or fluids isolated therefrom. Accordingly, compositions comprising a carrier and an amount of an antibody having specificity for the EHOC-1 polypeptide effective to block binding of naturally occurring ligands to the EHOC-1 polypeptides. A monoclonal antibody directed to an epitope of an EHOC-1

polypeptide molecule present on the surface of a cell and having an amino acid sequence substantially the same as an amino acid sequence for a cell surface epitope of an EHOC-1 polypeptide shown in SEQ ID NO: 2, can be useful for this purpose.

The invention provides a transgenic non-human mammal that is capable of expressing nucleic acids encoding an EHOC-1 polypeptide. Also provided is a transgenic non-human mammal capable of expressing nucleic acids encoding an EHOC-1 polypeptide so mutated as to be incapable of normal activity, i.e., does not express native EHOC-1. The present invention also provides a transgenic non-human mammal having a genome comprising antisense nucleic acids complementary to nucleic acids encoding an EHOC-1 polypeptide so placed as to be transcribed into antisense mRNA complementary to mRNA encoding an EHOC-1 polypeptide, which hybridizes thereto and, thereby, reduces the translation thereof. The nucleic acid may additionally comprise an inducible promoter and/or tissue specific regulatory elements, so that expression can be induced, or restricted to specific cell types. Examples of nucleic acids are DNA or cDNA having a coding sequence substantially the same as the coding sequence shown in SEQ ID NO: 1. An example of a non-human transgenic mammal is a transgenic mouse. Examples of tissue specificity-determining elements are the metallothionein promoter and the L7 promoter.

Animal model systems which elucidate the physiological and behavioral roles of EHOC-1 polypeptides are produced by creating transgenic animals in which the expression of the EHOC-1 polypeptide is altered using a variety of techniques. Examples of such techniques include the insertion of normal or mutant versions of nucleic acids encoding an EHOC-1 polypeptide by microinjection, retroviral infection or other means well known to those

skilled in the art, into appropriate fertilized embryos to produce a transgenic animal. (See, for example, Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual (Cold Spring Harbor Laboratory, 1986). Another technique, homologous recombination of mutant or normal versions of these genes with the native gene locus in transgenic animals, may be used to alter the regulation of expression or the structure of the EHOC-1 polypeptide (see, Capecchi et al., Science 244:1288 (1989); Zimmer et al., Nature 338:150 (1989); which are incorporated herein by reference). Homologous recombination techniques are well known in the art. Homologous recombination replaces the native (endogenous) gene with a recombinant or mutated gene to produce an animal that cannot express native (endogenous) protein but can express, for example, a mutated protein which results in altered expression of the EHOC-1 polypeptide. In contrast to homologous recombination, microinjection adds genes to the host genome, without removing host genes. Microinjection can produce a transgenic animal that is capable of expressing both endogenous and exogenous EHOC-1 protein. Inducible promoters can be linked to the coding region of the nucleic acids to provide a means to regulate expression of the transgene. Tissue specific regulatory elements can be linked to the coding region to permit tissue-specific expression of the transgene. Transgenic animal model systems are useful for in vivo screening of compounds for identification of specific ligands, i.e., agonists and antagonists, which activate or inhibit protein responses.

The nucleic acids, oligonucleotides (including antisense), vectors containing same, transformed host cells, polypeptides and combinations thereof, as well as antibodies of the present invention, can be used to screen compounds in vitro to determine whether a compound functions as a potential agonist or antagonist to the invention polypeptide. These in vitro screening assays

provide information regarding the function and activity of the invention polypeptide, which can lead to the identification and design of compounds that are capable of specific interaction with one or more types of polypeptides, peptides or proteins.

In accordance with still another embodiment of the present invention, there is provided a method for identifying compounds which bind to EHOC-1 polypeptides. The invention proteins may be employed in a competitive binding assay. Such an assay can accommodate the rapid screening of a large number of compounds to determine which compounds, if any, are capable of binding to EHOC-1 proteins. Subsequently, more detailed assays can be carried out with those compounds found to bind, to further determine whether such compounds act as modulators, agonists or antagonists of invention proteins.

In another embodiment of the invention, there is provided a bioassay for identifying compounds which modulate the activity of invention polypeptides. According to this method, invention polypeptides are contacted with an "unknown" or test substance (in the presence of a reporter gene construct when antagonist activity is tested), the activity of the polypeptide is monitored subsequent to the contact with the "unknown" or test substance, and those substances which cause the reporter gene construct to be expressed are identified as functional ligands for EHOC-1 polypeptides.

In accordance with another embodiment of the present invention, transformed host cells that recombinantly express invention polypeptides can be contacted with a test compound, and the modulating effect(s) thereof can then be evaluated by comparing the EHOC-1-mediated response (via reporter gene expression) in the presence and absence of test compound, or by comparing

the response of test cells or control cells (i.e., cells that do not express EHOC-1 polypeptides), to the presence of the compound.

As used herein, a compound or a signal that
5 "modulates the activity" of an invention polypeptide refers to a compound or a signal that alters the activity of EHOC-1 polypeptides so that the activity of the invention polypeptide is different in the presence of the compound or signal than in the absence of the compound or signal. In
10 particular, such compounds or signals include agonists and antagonists. An agonist encompasses a compound or a signal that activates EHOC-1 protein expression. Alternatively, an antagonist includes a compound or signal that interferes with EHOC-1 protein expression. Typically, the effect of
15 an antagonist is observed as a blocking of agonist-induced protein activation. Antagonists include competitive and non-competitive antagonists. A competitive antagonist (or competitive blocker) interacts with or near the site specific for agonist binding. A non-competitive antagonist
20 or blocker inactivates the function of the polypeptide by interacting with a site other than the agonist interaction site.

As understood by those of skill in the art, assay methods for identifying compounds that modulate EHOC-1
25 activity generally require comparison to a control. One type of a "control" is a cell or culture that is treated substantially the same as the test cell or test culture exposed to the compound, with the distinction that the "control" cell or culture is not exposed to the compound.
30 For example, in methods that use voltage clamp electrophysiological procedures, the same cell can be tested in the presence or absence of compound, by merely changing the external solution bathing the cell. Another type of "control" cell or culture may be a cell or culture
35 that is identical to the transfected cells, with the

exception that the "control" cell or culture do not express native proteins. Accordingly, the response of the transfected cell to compound is compared to the response (or lack thereof) of the "control" cell or culture to the same compound under the same reaction conditions.

In yet another embodiment of the present invention, the activation of EHOC-1 polypeptides can be modulated by contacting the polypeptides with an effective amount of at least one compound identified by the above-described bioassays.

The invention will now be described in greater detail with reference to the following non-limiting examples.

EXAMPLE 1

Construction of BAC Contig

BAC library construction of total human genomic DNA was described elsewhere Shizuya et al., *Proc. Natl. Acad. Sci. USA* 89:8794-8797 (1992). BAC clones were screened by PCR using STSs (PFKL, D21S25, D21S154, CD18). The loci of these BAC clones were confirmed by fluorescence *in-situ* hybridization. The sizes of inserts of these clones were measured by running pulsed-field gel electrophoresis after digesting DNA with NotI.

EXAMPLE 2

Direct cDNA Selection

Direct selection procedures were similar to those of Morgan et al. Morgan et al., *Nucleic Acid Res.* 20:5173-5179 (1992) with some modifications. Total RNA was isolated from 14 week trisomy 21 fetal brain using TRI region™ (Molecular Research Center, Inc.). Poly (A)⁺ RNA was isolated using Poly (A) Quick[®] mRNA isolation kit

(STRATAGENE). Double stranded cDNA was synthesized using SuperScript™ Choice System (GIBCO BRL) from 5µg trisomy 21 fetal brain poly (A)⁺ RNA using 1µg oligo (dT)₁₅ or 0.1µg random hexamer. The entire synthesis reaction was purified by Gene Clean[®] II kit (BIO101, Inc.) and was kinased. Sau3AI linker was attached to cDNA and digested with Sau3AI. The reaction was purified using Gene Clean. MboI linker I Morgan et al., *Nucleic Acid Res.* 20:5173-5179 (1992) was attached to the cDNA and purified by Gene Clean. Product was amplified by PCR using one strand of MboI linker (5'CCTGATGCTCGAGTGAATTC3') as a primer. Cycling conditions were 40 cycles of 94°C/15 seconds, 60°C/23 seconds, 72°C/2 minutes in a 100µl of 1xPCR buffer (Promega), 3mM MgCl₂, 5.0 units of Taq polymerase (Promega), 2µM primer and 0.2mM dNTPs. Five kinds of BAC DNA (total 2.5µg) was prepared using QIAGEN plasmid kit and was biotinylated using Nick Translation Kit and biotin-16-dUTP (Boehringer Manneheim). 3µg of heat denatured PCR amplified cDNA was annealed with 3µg of heat denatured COT1 DNA (BRL) in 100µl hybridization buffer (750mM NaCl, 50mM NaPO₄(pH7.2), 5mM EDTA, 5x Denhardt's, 0.05% SDS and 50% formamide) at 42°C for two hours. After prehybridization, heat denatured 1.2µg biotinylated BAC DNA was added and incubated at 42°C for 16 hours. cDNA-BAC DNA hybrid was precipitated with EtOH and dissolved to 60µl of 10mM Tris-HCl (pH 8.0), 1mM EDTA. After addition of 40µl 5M NaCl, the DNA was captured on magnetic beads (Dynabeads M-280, Dynal) at 25°C for 1 hour with gentle rotating. The beads were washed twice by pipetting in 400µl of 2x SSC, setting in magnet holder (MPC-E_{TM}, Dynal) for 30 seconds and removing supernatant. Four times additional washes were done in 0.2x SSC at 68°C for 10 minutes each with transferring beads to new tubes at each time. cDNAs were eluted in 100µl of distilled water for 10 minutes at 80°C with occasional mixing. The eluted cDNAs were amplified by PCR as described above. After repeating selection procedure on magnetic beads twice,

amplified cDNAs were digested with EcoRI and subcloned into pBluescript II.

EXAMPLE 3

Southern blot analysis

5 Gel electrophoresis of DNA was carried out on
0.8% agarose gels in 1xTBE. Transfer of nucleic acids to
Nybond N+ nylon membrane (Amersham) was performed by
following manufacturer's instruction. Probes were labelled
by RadPrime Labeling System (BRL). Hybridization was
10 carried out at 42°C for 16 hours in 50% formamide, 5xSSPE,
5x Denhardt's 0.1% SDS, 100µg/ml denatured salmon sperm
DNA. The filters were washed once in 1xSSC, 0.1% SDS at
room temperature for 20 minutes, twice in 0.1xSSC, 0.1% SDS
for 20 minutes at 65°C. Blots were exposed to X-ray films
15 (Kodak, X-OMAT-AR).

EXAMPLE 4

cDNA Library Screening

A trisomy 21 fetal brain cDNA library was
constructed using ZAP-cDNA[®] synthesis kit (STRATAGENE) which
20 generates unidirectional cDNA library. Briefly, double-
stranded cDNA was synthesized from 5µg trisomy 21 fetal
brain poly(A)⁺ RNA using a hybrid oligo(dT)-XhoI linker
primer with 5-methyl dCTP, attached EcoRI linker, digested
with EcoRI and XhoI, and cloned into UNI-ZAP XR vector.
25 The library was packaged using Gigapack[®] II Gold packaging
extract. The titer of the original library was 1.1×10^6
p.f.u./package. The library was amplified once. Blue-
white color assay indicated that 99% clones have inserts.
The average size of the inserts was 1.9 kb calculated from
30 14 clones.

The screening of trisomy 21 fetal brain cDNA
library was performed using selected cDNA fragments.

5

10

19

19

2

2

2

2

-

..